# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From - To)* |
|---|---|---|
| 28-02-2012 | Final Report | 01-05-2009 to 30-11-2011 |

**4. TITLE AND SUBTITLE**
ROBUST DYNAMIC VISION METHODS FOR PERSISTENT SURVEILLANCE AND ENHANCED VEHICLE AUTONOMY

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
FA9550-09-1-0253

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Mario Sznaier and Octavia I. Camps

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Northeastern University
360 Huntington Ave, Boston, MA, 02115

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Dynamics and Control Program, AFOSR.

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
AFRL-OSR-VA-TR-2012-0521

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
A

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**
This research addressed the USAF's unprecedented proactive persistent surveillance Long Term Challenge. Specifically, we aimed at a substantial enhancement of the ability to conduct autonomous, video based, persistent intelligent surveillance, reconnaissance and threat assessment in highly uncertain, adversarial scenarios such as urban environments. At its core was a novel approach, stressing dynamic models as key enablers for finding, tracking and anticipating/assessing behavior of multiple targets using as inputs data streams from spatially distributed sensors. It included both theory developments in an emerging new field –dynamics based extraction of information sparsely encoded in high dimensional data– and an investigation of implementation issues.

**15. SUBJECT TERMS**
Persistent Surveillance, Vision Based Tracking and Activity Monitoring, Hybrid Systems Identification and Model (In)Validation, Manifold Embedding

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| | | | | | Mario Sznaier |
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | UU | 9 | **19b. TELEPHONE NUMBER** *(include area code)* |
| U | U | U | | | 617 373 5364 |

ROBUST DYNAMIC VISION METHODS FOR PERSISTENT SURVEILLANCE AND ENHANCED VEHICLE AUTONOMY

Mario Sznaier and Octavia I. Camps
Department of Electrical and Comp. Engineering
Northeastern University
{msznaier, camps}@coe.neu.edu

FINAL REPORT

Grant/Contract Number FA9550-09-1-0253.

Dynamics and Control Program, Dr. Fariba Fahroo, Program Manager.

**Abstract**

This research addressed the *USAF's unprecedented proactive persistent surveillance* Long Term Challenge. Specifically, we aimed at a substantial enhancement of the ability to conduct autonomous, video based, persistent intelligent surveillance, reconnaissance and threat assessment in highly uncertain, adversarial scenarios such as urban environments. At its core was a novel approach, stressing dynamic models as key enablers for finding, tracking and anticipating/assessing behavior of multiple targets using as inputs data streams from spatially distributed sensors. It included both theory developments in an emerging new field –dynamics based extraction of information sparsely encoded in high dimensional data– and an investigation of implementation issues.

## Motivation

Controlled dynamic vision, the confluence of computer–vision and control, is positioned in an optimal situation to address the specific needs created by the move of the USAF towards a "digitized" battlefield. Smart UAVS can carry out intelligence gathering, target tracking and airspace denial, while minimizing the risk of loss of life. Proactive interfaces can interact better with human operators, allowing them to concentrate on critical tasks. In addition, the same technology can substantially benefit the general population. User aware environments can enable an aging population to carry on independent lives. Finally, intelligent surveillance systems capable of detecting *suspicious* activities will improve our ability to prevent tragedies.

Several proof–of–concept systems illustrating the ability of *dynamic* vision to successfully handle many of the challenges posed by these applications have already been developed. However, successful autonomous operation of these systems in highly uncertain, unstructured environments requires developing new mechanisms for robustly and timely extracting actionable information that is sparsely encoded in extremely large data streams.

**(a) (b) (c) (d)**

**Fig. 1: Examples of sparsely encoded visual information. (a) Target tracking in an urban canyon. (b) and (c) sample frames showing contextually abnormal events: onset of a tunnel fire and a person entering through an exit. (d) Tracking multiple targets. In all cases less than $O(10^{-6})$ of the data is relevant.**

The challenges entailed in this task are illustrated in Fig. 1. In all cases, decisions must be taken based on events discernible only in a small fraction of a very large data record: a short video sequence adds up to megabytes, yet actionable information (a change of behavior of a single target), may be encoded in just a few frames, e.g. less than $10^{-6}$ of the total data. Additional challenges arise from the quality of the data, often fragmented and corrupted by noise.

This research sought to address these issues by developing methods at the confluence of robust dynamical systems, information based complexity, machine learning and computer vision, laying the foundation for a new class of robust, autonomous vision–based systems.

### Description of the Approach and Results Obtained.

Below we summarize the results obtained in the course of this research. Technical details are provided in the cited publications, which can be obtained by contacting the authors or from http://robustsystems.ece.neu.edu. This site also contains presentations explaining these results in detail and several demos.

***Conceptual foundation: Dynamic models as information encoding and predictor paradigms***. The basic premise underlying this research was that relevant spatio/temporal information, at the granularity levels required by autonomous vision–based systems endowed with analysis and decision making capabilities, can be compactly encapsulated in dynamic models, whose rank, a measure of the dimension of useful information, is often far lower than the raw data dimension. This premise, amounting to a reasonable "localization" hypothesis for spatio/temporal correlations, allows for reducing each subproblem –tracking, multicamera coordination, dynamic data interpolation and segmentation, robust decision making– to the prototype system-theoretical problems discussed below. Embedding these problems in the conceptual world of dynamical systems, made available an extremely versatile ensemble of methods that allows for recasting them into a tractable, finite dimensional convex optimization form that can be efficiently solved.

**Basic Science Problems.** Application of the ideas outlined above to the problems arising in the context of persistent surveillance required addressing the following basic science issues:

**(a) Robust identification of Hammerstein-Wiener systems with high dimensional output spaces.** The problems of interest in this research were characterized by the need to identify systems whose outputs evolve in extremely high dimensional spaces: tracking target motion and appearance changes involve considering the evolution of $\approx O(10^3)$ pixels, even if using low resolution images. To address this challenge, we exploited the high degree of correlation between these pixels to embed the raw data in low dimensional manifolds. Since the projections to/from these manifolds can be modeled as memoryless non-linearities, this approach led to the identification problem shown in Fig. 2. Here $\Pi_i(.)$ and $\Pi_o(.)$ are memoryless nonlinearities, u, y, $u_m$ and $y_m$ represent the respective input, the raw data and their projection on the low dimensional manifold, and S is the dynamic model. In the course of this research we have established that these problems are generically NP-hard [P1] and proposed a polynomial time relaxation with suboptimality certificates [P2]. For the special case of rational nonlinearities (corresponding to rational embeddings), we have developed a computationally attractive relaxation based on recent results on polynomial optimization [P12,P22]. For the case of general nonlinearities, we have developed a convex optimization based approach for identifying the embedding manifold and correspondences by reducing the problem to a nuclear norm minimization subject to local isometric constraints. The resulting algorithm has the ability to exploit spatio-temporal dynamic constraints (while retaining computational complexity comparable to the state-of-the art) thus leading to embeddings that are robust to outliers and provide the most parsimonious model that explains the data [P19]. Finally, we have taken preliminary steps towards developing fast algorithms that exploit the intrinsic structure of the problem, leading to substantial improvement, both in computational time and memory requirements, over conventional semi-definite programming tools [P30].

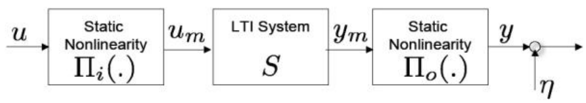



**Fig 2. Dynamic manifold learning as a Hammerstein-Wiener identification problem.**

**(b) Robust identification of hybrid systems.** Cases involving a transition between different models, e.g. substantially different appearances, can be modeled as a mode-transition in a piecewise affine switched system. In the case of noisy measurements, existing identification methods lead to computationally hard problems with poor scaling properties. We have shown that these difficulties can be circumvented by recasting the problem into a dynamic sparsification form, where one seeks the sparsest dynamical model that explains the data. Exploiting recent results from polynomial optimization allows for developing tractable relaxations with optimality certificates. Moreover, these relaxations exploit the underlying structure of the problem to substantially reduce the computational burden [P5,P8,P9,P13,P16,P17,P24,P26,P27,P28]. Finally, a feature that rendered the problems considered in this research challenging is the fact that the data records are often fragmented or corrupted, due to sensor or communication channel outages. As shown in [P21], this situation can be handled by introducing an additional set of variables, subject to structural sparsity constraints in the resulting optimization problems.

**(c) Robust model (in)validation of hybrid systems.** A crucial step before using the models identified in (b) above, is to check their validity against additional experimental data. A unique difficulty in validating hybrid models, is the fact that the mode signal is typically unmeasurable. As part of this research we have obtained a necessary and sufficient condition for a switched affine model to be (in)validated by the experimental data. The starting point is to recast the (in)validation problem as one of checking whether a semialgebraic set is empty. By using recent results on sparse polynomial optimization we have shown that this condition is equivalent to strict positivity of the solution of a related, convex optimization problem [P14,P24].

**(d) Constrained interpolation of noisy data.** In most surveillance scenarios only partial data is available, due for instance to occlusion or limited sensing/transmitting capabilities. In these situations, it is of interest to estimate the missing data, for instance in order to perform data association (e.g. stitch tracklets), or to uncover correlations mediated by the missing elements. We have shown that this interpolation can be reduced to a rank minimization problem, which in turn (due to its Hankel structure) can be efficiently solved using convex relaxations [P7,P11].

**(e) Robust estimation under $l^{\infty}$ bounded disturbances.** Traditional noise models often do not capture key features of the problems of interest here. As a simple example, noise in images should be bounded. While in principle this feature can be captured using truncated distributions, the resulting problems are computationally hard. To circumvent this difficulty we are developing a new framework for robust estimation in the presence of unknown-but-bounded noise. Using a concept similar to superstability leads to robust filters that can be synthesized by simply solving a linear programming problem [P4,P25]. A salient feature of this framework is that it explicitly allows for trading off filter complexity against worst-case estimation error. We have extended these results to a class of switched systems [P18].

**(f) Robust identification of sparsely interconnected networks.** The class of problems that motivated this research are characterized by complex systems composed by many interacting agents, each endowed with its own dynamics. In these cases a single lumped model (e.g. modeling a complex scenario with several adversarial teams using a single "statistical mechanics" motivated model) is often inadequate for scene analysis and trajectory prediction. Rather, what is needed is a model that captures both the individual dynamics and the dynamics of the interaction between agents. As part of this research we have shown that these models can be obtained by describing the interacting systems as a graph, where each node represents an homogeneous set of agents (with its own dynamical model), and the links, also dynamical models, account for the interactions amongst groups. As described in detail in [P23] the problem of identifying both the graph topology and the individual dynamical models can be reduced to a convex optimization problem (via group sparsity arguments) and efficiently solved by an algorithm that only uses local information.

**<u>Applications.</u>** The basic tools outlined above served as key enablers to address the following practical problems arising in the context of persistent surveillance:

**(g) Robust Tracking.** We have developed a new class of filters that do not require explicitly finding a model of the underlying process and have built-in adaptation capabilities. The main idea is to predict the next position of the target as the one that is maximally compatible with existing data, in the sense of leading to the minimum order interpolant. In turn, this problem can be recast into that of minimizing the rank of a suitable constructed Hankel matrix, and relaxed to a convex optimization using tools similar to those used in compressed sensing. The effectiveness of this approach is shown in Fig 3. In addition, in this context model

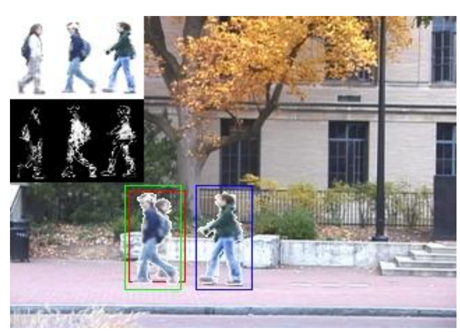
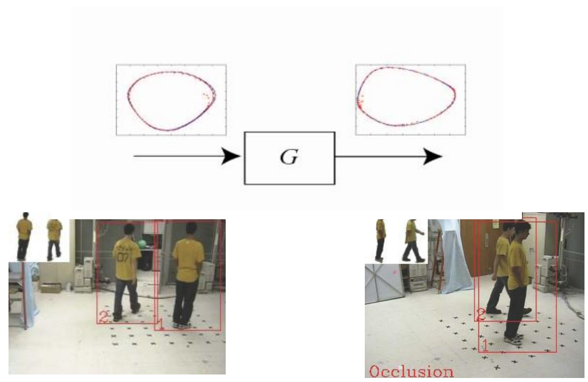


Fig. 3 Sustained tracking in the presence of occlusion.

switches are indicated by a sharp increase in the rank of the Hankel matrix, providing a computationally efficient way for segmenting high volume temporal data [P3,P7,P11].

**(h) Data Integration from Multiple Cameras.** In order for a multi-camera tracking system to take full advantage of the additional information available from its multiple sensors, it must maintain consistent identity labels of the targets across views and recover their 3D trajectories. We have developed a new approach to the problem of finding correspondences across views that does not require feature matching, camera calibration or planar assumptions. The key idea is to exploit the high spatio/temporal correlation between frames and across views by (i) associating to each viewpoint a set of intrinsic coordinates on a low dimensional manifold obtained using the identification methods described in (a) above, and (ii) finding an operator that maps the dynamic evolution of points over manifolds corresponding to different viewpoints. Then, correspondences can be found by simply running a sequence of frames observed from one view through the operator to *predict* the corresponding current frame in the other view (Fig. 4) [P3,P20].




**Left camera view** **Right camera view**

**Fig. 4 Using dynamic manifold mappings to recreate the appearance of an occluded person.**

**(i) Recovering 3D geometry from 2D data.** We have developed an efficient algorithm based upon recasting the problem into a Wiener system identification form. By exploiting dynamical information, this approach can recover the geometry of the scene up to an overall scaling constant. For comparison, existing approaches can recover scene information only up to a (time-varying) projective transformation that does not preserve Euclidian geometry [P6,P10].

**Fig. 5. Recovering the 3D geometry of a scene. Left: sample frame. Right: recovered geometry (red) superimposed on the ground truth (blue)**

**(j) Activity Recognition.** We have shown that this problem can be translated into a "behavioral" model invalidation form, where the goal is to establish whether two given time series are trajectories (or "behaviors") of the same underlying dynamical model. The resulting problem can be recast into a convex semidefinite program and efficiently solved [P15,P16,P29]. Applying these ideas to the problem of classifying activities from the challenging TV interactions database led to a 68% success rate, compared against the best reported performance in the literature of 54.5%. In the simpler case of single activities from the KTH database, the proposed approach had a 93.6% success rate, compared to 92.1% achieved by existing algorithms [P15].

**(k) Detecting Contextually Abnormal Events.** This problem fits naturally in the framework developed in this research by associating activities to an underlying dynamical model. In this context, a video sequence does not contain abnormal activities if an only if the observed data corresponds to an admissible trajectory of a system described by a graph, where each node corresponds to the dynamical system associated with a normal activity and links detecting abnormal events reduces to the hybrid model (in)validation problem discussed in item (c) above. A simple example illustrating these ideas is shown in Fig. 6. Further details are given in [P21].

**Fig. 6: Anomalous behavior detection as a switched (in)validation problem. The top sequence (walk– wait–walk) is not (in)validated since both activities are in the database. The bottom sequence(walk-jump) is flagged as abnormal since it cannot be generated by switching amongst models in the database.**

**(l) Detecting Coordinated Activities.** Causal correlations between individuals can be detected by reducing the problem to the sparse network identification problem described in (f). In this context, each node in the graph corresponds to the observed activity of a given agent, and each link indicates the presence of a causal correlation. It is worth emphasizing that this approach requires neither previous training nor repetitive activities. An example of application of these ideas is shown in Fig. 7, where they were used to identify the correlation between agents in a complex scenario. Further details are provided in [P23].

**Fig. 7: Sample frames from a doubles tennis game with identified causal connections superimposed.**


**Acknowledgment/Disclaimer**

This work was sponsored (in part) by the Air Force Office of Scientific Research, USAF, under grant/contract number FA9550-09-1-0253. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFOSR or the U.S. Government.

**Personnel Supported During Duration of Grant:**

| | |
|---|---|
| Octavia Camps | Professor, ECE Dept, Northeastern University. |
| Mario Sznaier | Dennis Picard Trustee Professor, ECE Dept, Northeastern University. |
| Necmiye Ozay | Graduate Student, Northeastern University (Ph.D granted 6/2010). |
| Mustafa Ayazoglu | Graduate Student, Northeastern University. (Ph..D expected 6/2012) |
| Xiong Fei | Graduate Student, Northeastern University. |
| Burak Yilmaz | Graduate Student, Northeastern University. |

**Publications:**

P1. M. Sznaier, *"Computational Complexity Analysis of Set Membership Identification of Hammerstein and Wiener Systems,"* Automatica, 45, 3, pp. 701 – 705, 2009.

P2. M. Sznaier, W. Ma, O. Camps and H. Lim, *"Risk Adjusted Set Membership Identification of Wiener Systems,"* IEEE Trans. Aut. Contr., 54, 5, 1147 – 1152, 2009.

P3. M. Sznaier and O. Camps, *"Motion Prediction for Continued Autonomy,"* in Encyclopedia of Complexity and Systems Science, R. A. Meyers, Editor-in-Chief, Springer, Berlin, 2009.

P4. F. Blanchini and M. Sznaier, *"A Convex Optimization Approach to Synthesizing Bounded Complexity $l^\infty$ Filters,"* Proc. 48th IEEE Conf. Dec. Control, Shanghai, China, Dec. 16-18, 2009, pp. 217-222.

P5. N. Ozay, C. Lagoa and M. Sznaier, *"Robust Identification of Switched Affine Systems via Moments–Based Convex Optimization,"* Proc. 48th IEEE Conf. Dec. Control, Shanghai, China, Dec. 16-18, 2009, pp. 4686-4691.

P6. M. Sznaier, M. Ayazoglu and O. Camps, *"Using Dynamics to Recover 3–Dimensional Euclidian Structure from 2–Dimensional Perspective Projections,"*, Proc. 48th IEEE Conf. Dec. Control, Shanghai, China, Dec. 16-18, 2009, pp. 2414-2419.

P7. M. Sznaier, O. Camps, N. Ozay, T. Ding, G. Tadmor and D. Brooks, *"The Role of Dynamics in Extracting Information Sparsely Encoded In High Dimensional Data Streams,"* Dynamics of Information Systems, Hirsch, M.J.; Pardalos, P.M.; Murphey, R. (Eds.), pp. 1-28, Springer Verlag, 2010 .

P8. C. Feng, C. Lagoa and M. Sznaier, *"Hybrid System Identification via Sparse Polynomial Optimization*," Proc. 2010 American Control Conference, June 2010.

P9. N. Ozay, M. Sznaier, O. Camps and C. Lagoa, *"GPCA with Denoising: A Moments-Based Convex Approach,"* 2010 IEEE Conf. Comp. Vision and Pattern Recognition (CVPR), pages 3209-3216, July 2010.

P10. M. Ayazoglu, M. Sznaier and O. Camps, *"Euclidean Structure Recovery from Motion in Perspective Image Sequences via Hankel Rank Minimization,"* 11th European Computer

Vision Conference, Springer-Verlag Lecture Notes in Computer and Information Sciences, Vol 6312, pp. 71-84, Sept. 2010.

P11. M. Sznaier and O. Camps, *"Dynamics-based Extraction of Information Sparsely Encoded In High Dimensional Data Streams,"* (lead tutorial paper), Proc. 2010 IEEE Multi-Conference on Systems and Control, pages 1234 – 1245, Sept. 2010.

P12. M. Ayazoglu, M. Sznaier, C. Lagoa and O. Camps, *"A Moments-Based Approach to Estimation and Data Interpolation for a Class of Wiener Systems,"* 2010 IEEE Conf. Decision and Control, pp. 5674-5680, Dec. 2010.

P13. C. Feng, C. Lagoa, N. Ozay and M. Sznaier, "*Hybrid System Identification: An SDP Approach,"* 2010 IEEE Conf. Decision and Control, pp. 1546-1552, Dec. 2010

P14. N. Ozay, M. Sznaier and C. Lagoa, *"Model (In)validation of Switched ARX Systems with Unknown Switches and its Application to Activity Monitoring,"* 2010 IEEE Conf. Decision and Control, pp. 7624-7630, Dec. 2010.

P15. B. Li, M. Ayazoglu, T. Mao, O. Camps and M. Sznaier, *"Activity Recognition using Dynamic Subspace Angles,"* 2011 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3193–3200.

P16. M. Sznaier and O. Camps, *"A Rank Minimization Approach to Trajectory (In)Validation,"* In Proc. 2011 American Control Conference, pp. 675–680.

P17. C. Feng, C. Lagoa and M. Sznaier, *"Identifying Stable Fixed Order Systems From Time and Frequency Response Data,"* In Proc. 2011 American Control Conference, pp. 1253–1259

P18. M. Sznaier, B. Yilmaz, and F. Blanchini, *"Bounded Complexity l-infinity Filters for Switched Systems."* In Proc. 2011 American Control Conference, pp. 2006–2011.

P19. X. Fei, O. Camps and M. Sznaier, *"Low Order Dynamics Embedding for High Dimensional Time Series,"* Proc. 2011 ICCV, pp. 2368 – 2374.

P20. M. Ayazoglu, B. Li, C. Dicle, M. Sznaier and O. Camps, *"Dynamic Subspace-Based Coordinated Multicamera Tracking,"* Proc. 2011 ICCV, pp. 2462 – 2469.

P21. N. Ozay and M. Sznaier, *"Hybrid System Identification with Faulty Measurements and Its Application to Activity Analysis,"* Proc. 2011 IEEE Conf. Dec. Control, pp. 5011-5016

P22. B. Yilmaz, M. Ayazoglu, M. Sznaier and C. Lagoa, *"Convex Relaxations for Robust Identification of Wiener Systems and Applications,"* Proc. 2011 IEEE Conf. Dec. Control, pp. 2812-2818.

P23. M. Ayazoglu, M. Sznaier and N. Ozay, *"Blind Identification of Sparse Dynamic Networks and Applications",* Proc 2011 IEEE Conf. Dec. and Control, pp. 2944-2950.

P24. C. Feng, N. Ozay, C. Lagoa and M. Sznaier, *"Identification and Model (In)validation of Switched ARX Systems: Moment-Based Approaches,"* in Linear Parameter Varying Systems: New Developments and Trends, P. L. dos Santos, C. Novara and D. Rivera (Eds),World Scientific Publishing, pp 348-380, Dec. 2011.

P25. F. Blanchini and M Sznaier, *"A Convex Optimization Approach to Synthesizing Bounded Complexity $l^\infty$ Filters,"* IEEE Trans. Aut. Contr, Vol 57, (1), pp. 216 – 221, 2012.

P26. N. Ozay, M. Sznaier, C. Lagoa and O. Camps, *"A Sparsification Approach to Set Membership Identification of Switched Affine Systems,"* IEEE Trans. Aut. Contr, to appear, March 2012.

P27. C. Feng, C. Lagoa and M. Sznaier, *"A Convex Approach to Generalized Fixed Order Interpolation,"* Proc. 2012 American Control Conference, to appear, June 2012.

P28. N. Ozay, C. Lagoa and M. Sznaier, *"Robust Identification of Switched Linear Systems with Bounded Number of Subsystems,"* submitted to Automatica.

P29. B. Li, O. Camps and M. Sznaier, *"Cross-view Activity Recognition using Hankelets,"* submitted to the 2012 IEEE Conf. Comp. Vision and Pattern Recognition.

P30. M. Ayazoglu, M. Sznaier and O. Camps, *"Fast Algorithms for Structured Robust Principal Component Analysis,"* submitted to the 2012 IEEE Conf. Comp. Vision and Pattern Recognition.

**Honors & Awards Received**

M. Sznaier selected as plenary speaker for the following international conferences:

- 2012 IFAC Symposium on System Identification (SYSID 2012)
- 2012 IFAC Symposium on Robust Control Design (RoCond 2012)
- 2012 IEEE Mediterranean Control Conference (MeD 2012)
- Second International Conference on the Dynamics of Information Systems (2010)
- First International Conference on the Dynamics of Information Systems (2009).

**AFRL Point of Contact**

Dr. Robert Murphey, Technical Advisor, Advanced Guidance Division, AFRL/RWG, Eglin AFB, FL 32542, Phone 850-882-2220.

New Discoveries**: None.**

**Transitions:** The target tracking and anomaly detection algorithms are currently being evaluated by the Federal Transportation Security Agency (TSA) for potential deployment at selected venues.